International Journal of Agricultural
Science and Research (IJASR)
ISSN(P): 2250-0057; ISSN(E): 2321-0087
Vol. 7, Issue 3, Jun 2017, 9-20


TRANS STELLAR
•Journal Publications •Research Consultancy

# GENETIC DIVERSITY, POPULATION STRUCTURE AND MARKER TRAIT ASSOCIATIONS FOR QUANTITATIVE AND QUALITATIVE TRAITS IN TOBACCO USING MICROSATELLITES MARKERS

**VN CHAKRAVARTHI DHAVALA, BNV PRIYA, CT GANESH,**
**U CHIDANAND, TV REDDY, DL RUDRESH & GVS SAIPRASAD**

*ITC - Life Sciences and Technology Centre, Peenya Industrial Area, Bengaluru, India*

***ABSTRACT***

***The current study was aimed to resolve the genetic diversity and relatedness of Nicotiana Spp. by using randomly distributed simple sequence repeats (SSR) loci across all chromosomes for exploration of marker and trait association. A set of 372 Nicotiana accessions were genotyped using 149 genome wide SSR markers to assess the molecular genetic diversity and genetic relatedness. The study has revealed a total of 1721 alleles with mean PIC value of 0.47 and the mean heterozygosity of 0.53 for various loci. The model based population structure analysis inferred seven distinct subgroups which were further confirmed by classical molecular genetic diversity analysis. AMOVA analysis has explained that 6% of variation was due to difference between groups and the remaining 94% variation could be attributed by difference within groups. The result of cluster method of neighbour joining tree also separated same seven groups like structure analysis revealing that this population panel is suitable for the association analysis. The General, Mixed Linear and Q + ML Models were used to detect associations between markers and traits by considering population structure and relatedness. A total of 56whole genomic SSR markers showed significant associations with fifteen out of fifty four traits. FDR value considered for associations was>0.05. Our analysis revealed, 18 markers associated with chemistry traits. Similarly, a total of 33 markers associated with morphometric characters and 5markersassociated significantly with both chemistry and morph metric traits.***

***KEYWORDS: Genetic Diversity; Population Structure; Association Mapping; Microsatellite Markers; Candidate Gene & Nicotiana Tabacum L***

**Received**: Mar 25 2017; **Accepted**: Apr 10, 2017; **Published**: Apr 17, 2017; **Paper Id.**: IJASRJUN20172

Original Article

## INTRODUCTION

*Nicotiana* spp. is one of the most economically important non-food crops that are widely cultivated worldwide for leaf (Moon et al. 2009). Beside leaves as its economic part, seeds contain 38% of non-edible oil which could be an appropriate substitute for diesel fuel (Giannelos et al. 2002).Tobacco leaf is not only used for making cigarettes but also used both in traditional and concurrent medicine in treating insect bites, cuts and tumours (Ahmad et al. 2014). More importantly, tobacco is an attractive green bioreactor proved to produce plant made vaccines, enzymes, immune modulatory molecules such as cytokines and high value pharmaceuticals (McCormick 2011; Phoolcharoen et al. 2011; Xiao et al. 2015). Despite the potential usage of tobacco in pharmaceutical and commercial production, limited cultivars exist with less harm associated and desirable traits. In order to develop such varieties, knowledge of the genetic diversity, relationships and population structure of the breeding materials is of fundamental importance for the crop improvement of breeding lines. Different types of

tobaccos are defined earlier depending on their usage, their origin of production, intended use in cigar (filler, binder and wrapper) and cigarette manufacturing, method of leaf curing (fire, air, sun, and flue cured tobacco), morphological and biochemical characters (aromatic flue cured, bright leaf tobacco, Burley, Turkish and oriental) (Edrisi et al. 2012; Lewis et al. 2007). However, for precise genetic manipulation of complex traits like, yield, flavour, quality and low levels of harm compounds etc., the genetic and molecular basis of target traits needs to be investigated thoroughly.

Genetic diversity studies on tobacco using molecular markers have been carried out by number workers earlier (Chen et al. 2007; Fricano et al. 2012; Ganesh et al. 2014; Sarala et al. 2008; Yang et al. 2005; Xiao et al. 2007; Xaio et al. 2009; Zhang et al. 2006; Zhang et al. 2008). Genetic analysis of this crop requires several critical steps that include accurate phenotyping for various qualitative and quantitative traits along with reliable and reproducible genotyping with sequence based markers rather than randomly amplified dominant markers. Simple sequence repeats (SSRs), also known as microsatellites, produce codominant, multi allelic, reproducible bands on amplification (Jones et al. 1997) and were developed by several groups for use in *Nicotiana* spp. (Ashkan et al. 2014; Bindler et al. 2011; Darvishzadeh et al. 2014; Fricano et al. 2012). SSR markers can resolve population structure and relatedness precisely because they are highly polymorphic markers and are reproducible for use in marker assisted selection programs (Xu et al. 2008). So far, very few molecular diversity studies involving SSR marker have investigated on a large collection of tobacco accessions (Ashkan et al. 2014; Darvishzadeh et al. 2014; Fricano et al. 2012). Therefore the current study was aimed to resolve molecular diversity of existing *Nicotiana* collections and population structure using genome wide distributed SSR markers and evaluate its use in an association mapping study.

Association mapping harnesses the genetic diversity of natural populations to potentially resolve complex trait variation to single genes or stretch of nucleotides (Zhu et al. 2008). Conventional linkage analysis with experimental population derived from a bi-parental cross provides pertinent information about traits that tends to be specific to the same or genetically related populations, while results from association mapping are more applicable to a much wider germplasm base. The ability to map QTLs in collections of breeding lines, landraces, or simples from natural populations has great potential for future trait improvement and line development within short span of time. Further association mapping has been found effective for mining new markers and has been used with all major crops including maize, rice, barley, tomato, wheat, sorghum, sugarcane, soybean, grape and melon. For estimating Q and K, multiallelic and co dominant microsatellites can be used since they are selectively neutral (Zhu et al. 2009). As compared with other marker system such as SNPs, SSRs are relatively new alleles and show higher rates of mutation (Matsuoka et al. 2002). Hence, the current study was aimed at evaluating the genetic variation, population structure and relatedness of tobacco using 149 SSR markers and to identify associations with quantitative and qualitative traits in a collection of 372 *Nicotiana* accessions.

## MATERIALS AND METHODS

A total of 372 *Nicotiana* accessions were valuated during 2011, 2012 and 2013 seasons at different tobacco growing regions viz., NLS (Northern Light Soils), KLS (Karnataka Light Soils), SBCS (Southern Block Cotton Soils) and SLS (Southern Light Soils).Data on morphometric traits viz., plant height, leaf area index, number of leaves, stem diameter, internodal length, days to 50% flowering, cured leaf yield (gms) per plant, cured leaf quality (1 to 5 scale), fill value & trichome number has been recorded along with basic chemistry data (Viz., Nicotine, chlorides and total sugars) including TSNA (Viz., NNN, NAT, NAB, NNK) and total alcohols. Data on 36 metabolites from cured leaf samples were also recorded.

Microsatellite markers distributed randomly across the chromosomes were retrieved from the *in house*whole genome sequence of tobacco. SSRs with a motif length of 14bp and above were selected for designing primers. Primer pairs flanking SSRs were selected using Primer3 software (http://frodo.wi.mit.edu.primer3/). The key parameters set for primer design were as follows: primer length 18-24 bp with 20 bp optimum; PCR product size 100-300 bp; optimum annealing temperature 54°C; GC content 35-60% with 50% as the optimum. The canonical name proposed for designed markers includes Lab [ITC LSTC, Bengaluru (ILB)], Species [*Nicotiana tabacum* (NT)], and type of marker [Microsatellite (m)] and serial #of marker. Hence the markers developed in this study were named Ilbntm. The fluorescent dyes were selected and labelled to the forward primer according to the Bindler et al.(2011).DNA was extracted from leaf tissue by grinding with liquid nitrogen using CTAB method. DNA was diluted to a final concentration of 30ng $\mu l^{-1}$ for enabling polymerase chain reactions. DNA amplification was carried out according to Ganesh et al. (2014) by using FAM, HEX or ROX labelled SSR primers. The PCR products were size separated by capillary electrophoresis using an ABI Prism 3170 DNA analyser. Alleles were scored using Peak Scanner 3.25 software according to manufactures instructions. Based on the height of the chromatogram peaks, the allele frequencies were scored for PCR products of various SSR markers.

Using the scored molecular data, genetic diversity parameters such as number of alleles per locus, allele frequency, heterozygosity and polymorphic information index (PIC) were estimated using the program POWERMARKER Ver3.25 (Liu et al. 2005). To assess genetic structure, model based and distance based approaches were used. Model based Bayesian approach was executed with Structure ver 2.3.4 software (Pritchard et al. 2000). Five independent runs were performed setting the number of subpopulation (K) from 1 to 10, burn in time and Markov Chain Monte Carlo (MCMC) replication number both to 100000, and a model for admixture and correlated allele frequencies. The K-value was determined by the log likelihood for each K; Ln P (D) = L(K). The optimum k value was determined by plotting the mean estimate of the log posterior probability of the data L(K) against the given K value. True number of subpopulation was identified using the maximal value of L(K) according to Evanno et al. (2005). Inferred ancestry estimates of individuals (Q-matrix) were derived for the selected subpopulation (Pritchard et al. 2000). The kinship coefficient was estimated in Tassel 2.1 software.

The genetic distance between accessions was estimated using Nei coefficient with bootstrap procedure of resampling across markers and individuals from allele frequencies. To determine the association among the accessions, un weighted pair group method with arithmetic mean (UPGMA) was implemented using DA R win. The presence of molecular variance within and between hierarchical populations was estimated by Analysis of Molecular Variance (AMOVA) by Gen Al Ex 6.5. Associations between markers and traits were analysed using the TASSEL software with Q, K and Q+MLM models. A false discovery rate adjusted probability value of 0.05 was used as the threshold for significance of SSR–trait associations (Benjamini et al. 1995).

## RESULTS

Enotyping of 372 *Nicotiana* accessions using 149 SSR (Microsatellite) markers produced a total of 1721 alleles. The number of alleles per loci varied from 2 to 41 with an average of 11.55 alleles per locus. The highest number of 41 alleles were detected for the loci Ilbntm-172 followed by 34 in loci 3384 Ilbntm-175 and the lowest of 2 alleles was detected in two markers viz.,Ilbntm-225 and Ilbntm-487.However, we also identified the number of specific alleles across the germplasms and polymorphic alleles were used for the analysis (Figure 1). The average PIC value was found to be

0.474 ranging from a minimum of 0.101 to a maximum of 0.856. Expected heterozygosity or gene diversity (He) as computed is varied from 0.106 to 0.87 with an average of 0.534 (table S1).

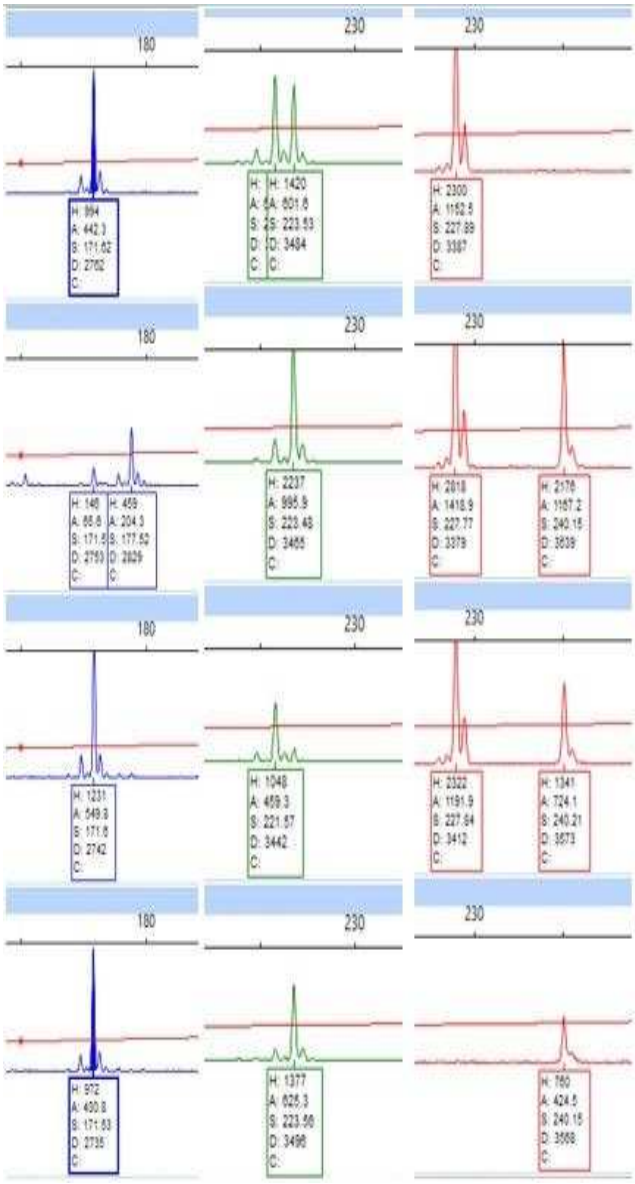


**Figure 1: Amplification Profile of Specifc Polymorphic Alleles across the Germplasms**

Population structure of the 372 accessions was analysed by Bayesian based approach. The estimated membership fractions of 372 accessions for different values of K ranged from 1 to 10 based on the distribution of 1757 alleles. STRUCTURE simulation demonstrated that the K value showed a modest peak at k=7, suggesting that seven sub-populations could contain all individuals with greatest probability (Figure 2). Hence a K value of 7 groups (subpopulations) was selected to describe the genetic structure of 372 accessions analyzed. Based on the membership fractions, the accessions with the probability of ≥80% were assigned to corresponding subgroups with others categorized as admixture. Cluster analysis based on Un weighted Pair Group Method with A thematic Mean (UPGMA) method using DAR win separated the accessions into seven main groups which showed similar result as STRUCTURE analysis.

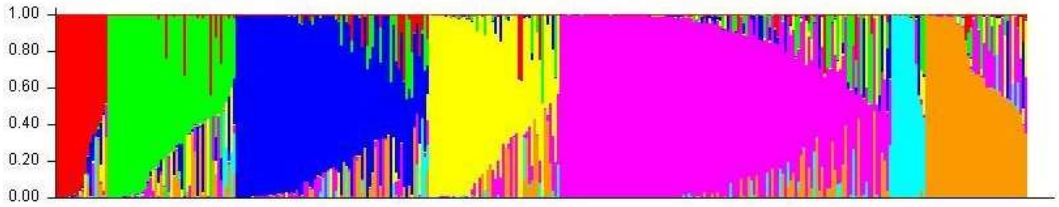


**Figure 2: Population Structure of Tobacco Accessions Based on 149 SSR Loci**

In order to evaluate genetic variation within and among group components, AMOVA (Analysis of Molecular Variation) was performed. This has revealed that 6% ($p>0.001$) of variation due to difference among groups and 25% ($p>0.001$) of variation due to difference among the individuals and the remaining 69% ($p>0.001$) variation could be attributed due to difference within individuals in the groups (table 1).

**Table 1: AMOVA Analysis: Partition of Genetic Variance between Groups, Among and Within the Individuals**

| Source of Variation | Df | Sum of Squares | Estimated Variance | % Variance | P Value |
|---|---|---|---|---|---|
| **Among Pops** | 6 | 1853.56 | 2.452 | 6% | 0.001 |
| **Among Indiv** | 365 | 24469.85 | 28.481 | 69% | 0.001 |
| **Within Indiv** | 372 | 3749.00 | 10.078 | 25% | 0.001 |
| **Total** | **743** | **30072.41** | **41.011** | | |

Each cluster distinguishes the genotypes clearly from the other. Subgroup1 comprised of 20 wilds and its relatives. Among these 10 originated from Maryland (*N. obtusifolia*,*N. alata*, *N. attennata*, *N. benavidesic*, *N. benthamiana*, *N. plumbaginifolia*, *N. rotundifolia*, *N. stocktonic*, *N. suvaveolence and N. motitina*), 4 from Venezuela, 2 from Columbia and rest were obtained from Ethiopia, Taiwan, Campania and New Guiana respectively. Subgroup2 consisted of 49 accessions comprising of cigar fillers, primitive FCV (3) lines and 13 cultivars of FCVs from Venezuela, 8 lines from Mexico, 6 lines from Colombia, 3 from Honduras and 3 from Maryland, 3 from North Carolina, 2 from Argentina, 2 from Brazil and rest are from Germany, Australia, Peru, South Carolina, Morza and Ecuador regions respectively. Majority of these lines are primitive FCV lines and rest admixed from other clusters ancestry. The subgroup3 congregated with 74 accessions. Of these, majority are air cured types mostly of *N. tabacum*var., Burley and admixtures of FCV, originated from China, Columbia, Maryland, North Carolina, Canada, Rhodesia, South Carolina, Venezuela Virginia Zambia and Zimbabwe. Whereas subgroup4 comprised of *N. tabacum*cultivarsViz., dark varieties, cigar wrapped and filler varieties obtained mostly from Brazil, Columbia, Maryland, Venezuela. The subgroup5 was dominated by *N. tabacum*cultivars consisting of 127 accessions of FCV. Of these, most of the lines originated from Maryland, North Carolina, South Carolina, Taiwan, Columbia, Virginia and rest were from different parts of the world. This group comprised of FCV commercial cultivars like Bottom Special, Cabbage, Cash, Gold Dollar, Mammoth Gold, Oxford, Silky Leaf, Silver Dollar, Virginia Gold, Mc Nair, Vista, Gold Leaf, K326, Sixie Bright, Coker, TT5, Cherry Red, Little Gold 1025, Delcrest, Griffin Special, Lonibow, Bonanza and Delhi 61 and other commercial varieties from different parts of the world. Sub population 6 consisted of 13 accessions, comprising of Oriental, macrophyllas and primitive types of tobacco. These lines mainly originated from Ecuador, Honduras, Venezuela, China and Peru. Subgroup7 comprised of Cigar wrappers, fillers and also some FCV cultivars obtained from Guyana, Italy, Mexico, Poland, South Africa, Soviet Union, Spain and other parts of the world. It has been observed that grouping and sub-grouping of the accessions within the different cluster was in accordance with their genealogies, origin and ancestral mating (Figure 3).

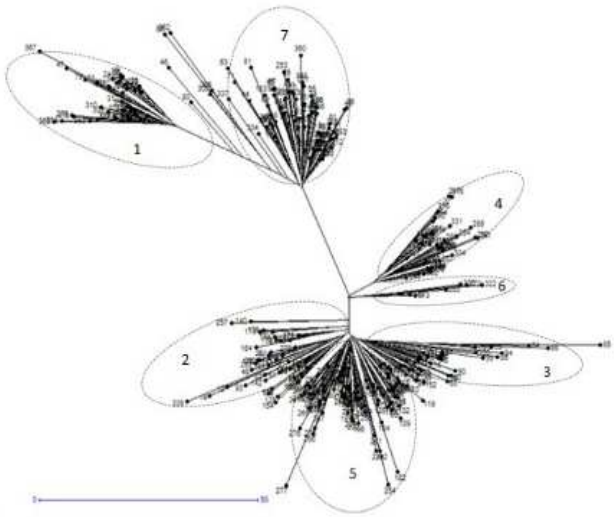


**Figure 3: Unrooted Neighbour Joining Tree of 372 Germplasms Based on Neis Genetic Distance**

Extensive phenotypic variations were also observed for all the measured quantitative traits in this collection, as shown by the descriptive statistics (tableS2). Correlation coefficient analysis was conducted for each trait from different seasons of 2011, 2012 and 2013 showed significant correlations at *P>0.01* level for all morphometric traits, suggesting the strong genetic impact than environmental influence (table S3). In addition to the existence of phenotypic variation in the traits and strong genetic impact, the average number of alleles per locus and genetic diversity has shown the existence of broad genetic base in this collection. The result of structure analysis is also in accordance with clustering method of neighbour joining tree, revealing that this population panel is suitable for the association analysis.

## ASSOCIATION MAPPING

The General and Mixed Linear Model and Q + MLM models were used to detect associations between markers and traits independently for 3 experimental seasons and 4 locations. This has revealed a total of 56 microsatellite marker associations with 15 traits out of54 traits studied. Q+MLM model showed least deviation of observed P-values from expected P-values in Q-Q plot when compared with that of Q (population structure) or K (kinship) model. Though this study showed many associations, only markers showed significant association across the seasons, locations and three analysis models were considered. FDR value considered for associations was>0.05. In particular, our analysis revealed 33microsatellite markers associated significantly with six morphometric characters Viz., plant height, number of leaves, leaf area, inter nodal length, stem diameter and curing quality. Four markers associated with nicotine and five markers associated significantly with total sugars, two markers associated with chlorides, and eleven with tobacco-specific nitrosamines viz., N-nitrosonornicotine (NNN), Nicotine-derived nitrosamine ketone (NNK), N-nitrosoanabasine (NAB) and N-nitrosoanatabine (NAT). Similarly, fifteen markers associated significantly with four megastigmasterone metabolites and four were associated with Hydroxy β-Damascone compound (table2). An attempt was also made simultaneously for cross validation of these markers using small set of separate FCV mixed population obtained from tobacco germplasm collection centre of NCSU, USA varying for few morphometric traits. The analysis carried out by same experimental models which revealed similar associations for 3 of the markers Viz., Ilbntm-478associated with the leaf area, marker Ilbntm-144with number of leaves and marker Ilbntm-479associated significantly with internodal length (p>0.05).

Majority of markers associated significantly with different traits are located near or at the structural and functionally controlling regions. Particularly,14 markers Viz., Ilbntm-61, Ilbntm-66, Ilbntm-78, Ilbntm-102, Ilbntm-106, Ilbntm-138, Ilbntm-142, Ilbntm-144, Ilbntm-154, Ilbntm-164, Ilbntm-242, Ilbntm-445, Ilbntm-478 and Ilbntm-479 located near or at the genic regions and 9 markers Viz., Ilbntm-14, Ilbntm-55, Ilbntm-81, Ilbntm-92, Ilbntm-123, Ilbntm-165, Ilbntm-193, Ilbntm-216 and Ilbntm-326 lies near or at the regulatory regions. Genes or regulatory regions found near the associated microsatellites are potentially related to the trait under consideration (Table2). Majority of the associated markers located near or at gene controlling regions revealed that these genes might be involved in the different biosynthetic pathways of the traits analysed or might be implied in their transcriptional regulation.

## DISCUSSIONS

The success of the association mapping depends upon the marker types, alleles affecting the expression of phenotypic traits and methods used for the marker traits association (Stich et al. 2005). In the current study, we have analyzed 220 *in-house* microsatellites distributed randomly across the whole genome of tobacco and identified a robust set of 149 SSRs that are polymorphic with Indian and exotic collections of tobacco accessions (Figure 1). Generally, SSRs have been well integrated with QTL mapping and Marker assisted selection (MAS) research in many commercial crops. SSRs are amenable to high throughput technologies, as compared to the ease of use of single nucleotide polymorphisms. SSRs have the unique advantage of being highly polymorphic and multiallelic in nature and are present in mutational hot spots in genomes (Shamjanaet al. 2015).

The mean number of alleles observed for each locus was 11.55 ranging from 2 to 41 alleles for marker. Similarly, the average PIC value was found to be 0.47 ranging from a minimum of 0.101 to a maximum of 0.856 with the expected average heterozygosity of 0.534 with a range of 0.106 to 0.87. These values are considerably higher than those reported in similar investigations carried out on tobacco accessions (Fricano et al. 2012; Moon et al. 2008; Moon et al. 2009; Sarala et al. 2008). The concordance of the STRUCTURE analysis results revealed the relationships among accessions were distributed over seven groups Viz., Wilds, Cigar, Burley, Dark, FCV, Oriental and others. The grouping and sub-grouping of the accessions within the different cluster was basically observed in accordance with their genealogies, origin and ancestral mating. Similar to that of model-based method, cluster analysis also separated these groups successfully and further confirms that these SSR markers were scattered over different parts of the genome and were grouped into seven different clusters according to their genealogies.

The distinct and homogeneous clustering of FCV, Burley, Oriental, Dark and Cigar obtained in our study is in agreement with Ashkanet al. (2014) and Fricano et al. (2012) where they observed homogeneous separate clusters for Oriental, FCV, Burley and cigar tobaccos because the most outstanding tobacco types, is most likely due to long years of selection in Europe/Middle East for the oriental types (Wolf et al. 1948), and to the adoption of a stringent conservative breeding strategies for FCV tobacco (Murphy et al. 1987).Fricano et al. (2012) also quoted that the genetic variability in FCV decreased significantly with the adoption of an advanced cycle pedigree breeding by using elite materials to produce breeding crosses. This could be the one main reason to congregate some of the FCVs into other groups in the present study. Compared to the Wilds, FCV, Oriental and Burley other clades Viz., Dark and Cigar showed more heterogeneity with admix of FCV and Primitive FCVs indicating the close relation of these accessions for FCV breeding. A set of 15 homogeneous filler and wrapped types of tobacco grouped with the heterogeneous admix of 24 FCV lines. Hence were named as others in the present study (Figure 2).

The clear separation among tobacco accessions as Wild, Cigar, Burley, Dark, Oriental, FCV and others observed in this study can explain the efficient use of microsatellite markers for the further association analysis. Our current finding is different from our early results of Ganesh et al. (2014), which identified narrow clustering of genotypes with limited divergence. This difference could attribute to the included additional germplasm collections. When these structured groups were used as a grouping factor to analyse the variance, high contribution of genetic variance was observed due to variation among and within samples at significant level p > 0.001 with 25 and 69% of variations respectively. However, the effect of groups contributing the variance was also found to be highly significant ($P$>0.001), revealing that the effect of groups on the partitioning of total genetic variance is equal to that of individuals. Data from this genetic diversity and population structure analysis revealed that this association panel showed a diverse genetic variation and therefore, could be used for the association analysis.

Correction for the confounding effects of population structure present in plant populations is essential for association mapping because the complex population structure may cause spurious correlations, which finally result in an elevated false-positive rate (Pritchard et al. 2000). To reduce the probability of detecting false positive marker-trait associations, one major method, the structured association (Pritchard et al. 2000 and Zhu et al. 2009), has been suggested to account for population structure. In this method, the Q matrix estimated by the program structure using a set of random markers is commonly incorporated in a General Linear Model (GLM) to test associations. However, Q matrix may not completely represent the population structure, although it can efficiently reduce the spurious associations. Yang et al. (2010) reported that structure program divides the panel into a few discrete populations, and the Q matrix only provides a rough dissection of population differentiation. Consequently, the K matrix Yu et al. (2006) calculated for familial relatedness has been broadened to combine with the Q matrix in a mixed linear model to improve the false positive detection rate, as described by Yu et al. (2006). Additional studies have demonstrated that the Q+K model controlling for population structure and genetic relatedness, is better than the Q and K models alone (Al-Maskriet al. 2012; Ahamad et al. 2014; Darvishzadeh et al. 2014; Shao et al. 2011; Yang et al. 2010; Yu et al. 2006). The present results are in agreement with these findings, and thus we considered resulted reduced number of significant markers and coefficients of determination across the seasons and locations using Q, K and Q+K analysis models. Differences in results among the models has been explained earlier by Achleitner et al. (2008), in oats, and in tobacco Ahmad et al. (2014) and Darvishzadeh et al. (2014) stated that the these differences illustrate relative importance of different parts of the population structure accounted for by different models and combination of both Q and K provide the strongest reduction in coefficients of determination and presumably the best correction for population structure. Total 222 significant associations (P< 0.05) noticed with 56 markers on 54 different traits. Using GLM model we observed 31% of associations are significant at P<0.05 level and 61% at P<0.01 level, Using MLM model 25% at P<0.05 level and 57% at P<0.01. It is 23% at P<0.05 level and 59% at P<0.01 was observed by Q+MLM model. However in the current study, markers significant across the all models were considered. Coefficient of variation ranged from 2 to 97 with the highest noticed on nitrosamine ketone with markers Ilbntm-91and followed by Ilbntm-142and in quantitative traits, the highest of 40% variation observed on number of leaves with marker Ilbntm-91. Based on the results of the three models, there are 56 markers that can be considered to be the most interesting putative candidates for further study. However, we have also conducted concurrent cross validation using small set of separate FCV population varying for few morphometric traits using same models which in turn has revealed 3 markers Ilbntm-478, Ilbntm-144and Ilbntm-479 are consistently showingsignificant association with the leaf area, number of leaves and internodal length respectively.

It has been observed that some of the SSR markers used in the present investigation were similarly located near to or exactly on the QTLs described previously by others in their linkage mapping studies. For example, among the traits investigated, plant height is one of the important agronomic traits associated with Ilbntm-236showed synteny with 12$^{th}$ chromosome at 23cM distance of Bindler et al. (2011) is near to published major QTL-qPh12. Other Markers, Ilbntm-242 associated for internodal length in our study, showed synteny to QTL-qIL12 located at 15 cM on 12$^{th}$ chromosome and Ilbntm-130 associated with leaf area is syntenic to the QTL-qLWL6 on 6$^{th}$linkage group of Cheng et al. (2015), Similarly,Ilbntm-144 associated with number of leaves near to QTL-qLN7 and Ilbntm-171 associated with stem diameter near to QTL-qSG8-2 identified by Tong et al. (2012). Distance between associated syntenic markers and published QTLs is ranging from 15 cM to 50 cM revealing that these markers can be considered as putative candidates for associated traits.

Candidate genes found near the associated microsatellites are potentially related to the trait under consideration. Particularly, most of the associated markers located near or at gene controlling regions revealed that these genes might be involved in the different biosynthetic pathways of the traits analysed or might be implied in their transcriptional regulation. Majority of markers interacted with sequence coding for cell cycle signalling precursors indicate the direct role on morphological developments. Markers like Ilbntm-61, Ilbntm-66, Ilbntm-78, Ilbntm-102, Ilbntm-106, Ilbntm-138, Ilbntm-142, Ilbntm-144, Ilbntm-154, Ilbntm-164, Ilbntm-242, Ilbntm-445, Ilbntm-478 and Ilbntm-479 located near and majority are within the genic regions revealing the role for trait expression. Markers Ilbntm-14, Ilbntm-55, Ilbntm-81, Ilbntm-92, Ilbntm-123, Ilbntm-165, Ilbntm-193, Ilbntm-216 and Ilbntm-326 lies near or at the regulatory regions like Zinc finger protein, NIP-type end nuclease, WD Repeat, protein SEC13 homolog, Leucine rich repeat receptor, replication factor C37-kDa subunit, origin recognition complex subunit 3-like regulatory regions indicating that these regions might be involved in regulation of expression of the trait. In tobacco, curing quality is an important trait that associated significantly with marker Ilbntm-138which in turn interacted very closely with APC controlling protein regulatory region. APC/MPF regulation specifically is crucial for cell-cycle progression during the cell proliferation phase of leaf growth and apoptosis during the cell maturation.

## CONCLUSIONS

Overall, current study demonstrated significant levels of associations for the quantitative and some quality related traits of *Nicotiana* species. High genetic diversity association panel used in this study attributes the possibilities for improvement of quantitative and qualitative traits. With the combination of the Q matrix and kinship model many loci were detected consistently for the seasons and locations and also coincide with known major genes or QTLs, some marker –trait loci were not reported previously indicating the power of the association panel. The role of these regions needs to be further investigated. Additionally potential novel loci were identified that may help to better understand the architecture of complex genetic traits. However, Markers found annotated in this study needs to be validate further for implementation in Marker Assisted Selection in future either with bi-parental mapping populations or NAM RILs.

## ACKNOWLEDGEMENTS

The authors thank Dr. C.C. Lakshmanan, Chief Scientist –Research & technology innovation and Head of R&D, ITC-LSTC, ITC Ltd, Bangalore for his consistent support and suggestions during the study. Authors also thankful to Dr. Chitranjan Dar and Dr. T. Ramaswamy for consistent support. We are also thank Crop Development teams of ILTD at NLS and KLS for supporting us to conduct field experiments. We are also thankful to Dr. M. Mani, Dr. Carol Jacob Gem, Mr. P. Uma Mahesh for chemical analysis and Dr. Rajkumar and his team for their help during analysis. Editorial and

useful comments from the Corp R&D in-house manuscript committee are appreciated. Overall support from team Agrisciences, LSTC, Bengaluru is greatly acknowledged.

***REFERENCES***


1. *Achleitner A., Nichola AT., Zechner E and Buerstmayr H 2008 Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits Theor. Appl. Genet.* ***117,****1041-1053*
2. *Al-Maskri AY., Sajjad M and Khan SH 2012 Association mapping, a step forward to discovering new alleles for crop improvement. Int J Agric Biol.* ***14,*** *153-160*
3. *Ashkan B., Hamid HM., Darvishzadeh R and Farhad G 2014 Mixed linear model association mapping for low chloride accumulation rate in oriental-type tobacco (Nicotiana tabaccum L.) germplasm J. of Plant Intera.****9****, 666–672*
4. *Ahmad RD., Hossein S., Ghasem MN., Atefeh S and Mardavij SD 2014 Association analysis., genetic diversity and structure analysis of tobacco based on AFLP markers Mol Biol Rep.****41,*** *3317–3329*
5. *Benjamini Y and Hochberg Y 1995 Controlling the false discovery rate, A practical and powerful approach to multiple testing. J. R. Stat. Soc.* ***57,*** *289–300*
6. *Bindler G., Plieske J., Bakaher N., Irfan G., Ivanov N., Hoeven RVD., Ganal M and Donini P 2011 A high density genetic map of tobacco (Nicotiana tabacum L.) obtained from large scale microsatellite marker development. Theor Appl Genet.****123,*** *219-230*
7. *Cheng L., Yang A., Jiang C., Ren M., Zhang Y., Fen Q., Wang S., Guan Y and Luo C 2015 Quantitative trait loci mapping for plant height in tobacco using linkage and association mapping methods. Crop science.****55,*** *641-647*
8. *Chen XJ., Yang BC., Xiao BG and Shi CH 2007 Assessing the genetic diversity of tobacco germplasm using inter simple sequence repeat and inter retrotransposon amplification polymorphism markers. Ann Appl Biol.****150,*** *393-401*
9. *Darvishzadeh R.,Leila M., Hamid H M., Hernan L and Seyyed RA 2014Genetic variation in oriental tobacco (Nicotiana tabacum L.) by agromorphological traits and simple sequence repeat markers Revista Ciencia Agron.* ***44,*** *347-355*
10. *Edrisi MK., Samizadeh LH and Shoaei MD 2012 Assessing the genetic diversity of tobacco (Nicotiana tabacum L.) varieties. Crop Breeding Journal.****2,*** *125-132*
11. *Evanno G., Regnaut S and Goudet J 2005 Detecting the number of clusters of individuals using the software STRUCTURE, a simulation study. Mol Ecol.****14,*** *2611-2620*
12. *Fricano A., Nicolas B., Marcello DC., Pietro P., Paolo D., Alessandra S., Nikolai VI and Carlo P 2012 Molecular diversity., population structure., and linkage disequilibrium in a worldwide collection of tobacco (Nicotiana tabacum L.) germplasm. BMC Genetics.****2156,****13-18*
13. *Ganesh CT., Saiprasad GVS., Mohan RB., Sheshayee MS and Udayakumar M 2014 Genetic structure in FCV Tobacco population as assessed by multi locus genotyping using SSR Markers. Adv Crop Sci Tech****. 2,*** *1-9*
14. *Giannelos PN., Zannikos F., Stournas S., Lois E and Anastopoulos G 2002 Tobacco seed oil as an alternative diesel fuel, physical and chemical properties. Ind Crops Prod.* ***16,****1-9*
15. *Jones N., Helen O and Howard T (1997) Markers and mapping, we are all geneticists now. New Phytol.* ***137,*** *165-177*
16. *Liu KJ and Muse SV 2005 PowerMarker, integrated analysis environment for genetic marker data. Bioinformatics****21,*** *2128-2129*
17. *Lewis RS., Nicholson JS 2007Aspects of the evolution of Nicotiana tabacum L.., and the status of the United States Nicotiana*

*Germplasm Collection. Genet Res Crop Evol.**54,**727–740*

18. *McCormick AA 2011 Tobacco derived cancer vaccines for non-Hodgkin's lymphoma, perspectives and progress. Hum. Vaccin. **7,** 305-312*
19. *Moon HS., Nicholson JS and Lewis RS2008 Use of transferable Nicotiana tabacum L. microsatellite markers for investigating genetic diversity in the genus Nicotiana. Genome**51,**547–559*
20. *Moon HS., Nifong JM., Nicholson JS., Heineman A., Lion K., Hoeven RVD., Hayes AJ and Lewis RS 2009 Microsatellite-based analysis of tobacco (Nicotiana tabacumL.) genetic resources. Crop Science **49,** 2149-2159*
21. *Matsuoka Y., Mitchell S E., Kresovich S., Goodman M and Doebley J 2002 Microsatellites in Zea-variability., patterns of mutations and use for evolutionary studies. Theor. Appl. Genet.**104,**436-450*
22. *Murphy JP., Cox TS., Rufty RC and Rodgers DM 1987 A representation of the pedigree relationships among flue cured tobacco cultivars. Tobacco sci.**31,** 70-75*
23. *Phoolcharoen W., Bhoo SH., Lai H., Ma J., Arntzen CJ., Chen Q and Mason HS 2011 Expression of an immunogenic Ebola immune complex in Nicotiana benthamiana L. Plant Biotechnol. J. **9,** 807–816*
24. *Pritchard JK., Stephens M and Donnelly P 2000 Inference of population structure using multilocus genotype data. Genetics.**155,** 945-959*
25. *Sarala K and Rao RVS 2008 Genetic diversity in Indian FCV and Burly tobacco cultivars. J. Genetics. **87,** 159- 164*
26. *Shamjana U., Bharadwaj T and Grace T 2015 Microsatellites, A versatile marker for genetic evolutionary ecological studies. Int Jour Adv Biol Res.**5,** 86-95*
27. *Shao YF., Jin L., Zhang G., Lu Y., Shen Y and Bao JS 2011 Association mapping of grain color., phenolic content., flavonoid content and antioxidant capacity in dehulled rice. Theor Appl Genet. **122,**1005-1016*
28. *Stich B., Albrecht E., Frisch MM., Maurer HP., Heckenberger M and Reif JC 2005 Linkage disequibrium in European elite maize germplasm investigated with SSRs.Theo. Appl. Genet. **111,**723-730*
29. *Tong ZJ., Jiao FC., Wu X., Wang FQ., Chen XJ., Li XY., Gao YL., Zhang YH., Xiao BG and Wu WR (2012) Mapping of Quantitative Trait Loci Underlying Six Agronomic Traits in Flue-Cured Tobacco (Nicotiana tabacumL.) Agronomica Sinica.**38,** 1407−1415*
30. *Wolf A and Wolf FT 1948 The origin of tobaccos of the oriental type. Bull Torrey Bot Club.**75,** 51-55*
31. *Xiao ZL., Chuan SH., Yu MY Han YZ and Czech JG 2009 Genetic diversity among flue-cured tobacco cultivars on the basis of AFLP markers. Plant Breed.**45,** 155–159*
32. *Xiao BG., Yang BC 2007 Assessment of genetic diversity among tobacco germplasms by ISSR markers. Sci. Agric. Sinica.**40,** 2153-2161*
33. *Xaio B., Tan Y., Long N., Chen X., Tong Z., Dong Y and Li Y 2015 SNP based genetic linkage map of tobacco (Nicotiana tabacum L.) using next generation RAD sequencing. J. Bio Res.**22,** 1-11.*
34. *Xu Y and Jonathan HC 2008 Marker-Assisted Selection in Plant Breeding, From Publications to Practice. Crop Science.**48,** 391-407*
35. *Yang., BC., Xiao BG., Chen XJ and Shi CH 2005. Genetic diversity of flue-cured tobacco varieties band on ISSR markers. Hereditas.**27,**753-775*
36. *Yang XH., Yan JB., Shah T., Warburton ML., Li Q., Li L., Gao YF., Chai YC., Fu ZY., Zhou Y., Xu ST., Bai GH., Meng YJ.,*

*Zheng YP and Li JS 2010 Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet.**121,** 417-431*

37. *Yu JM., Pressoir G., Briggs WH., Bi IV., Yamasaki M., Doebley JF., McMullen MD., Gaut BS., Nielsen DM., Holland JB., Kresovich S., Buckler ES 2006 A unified mixed model method for association mapping that accounts for multiple levels of relatedness. Nat Genet.**38,** 203-208*
38. *Zhang HY., Liu XZ., He CS and Yang YM 2008 Genetic diversity among flue-cured tobacco based on RAPD and AFLP markers. Braz Arch Biol Technol.**51,**1097–1101*
39. *Zhang HY., Liu XZ., Li TS and Yang YM 2006 Genetic diversity among flue-cured tobacco (Nicotiana tabacum L.) revealed by amplified fragment length polymorphism. Bot Stud.**47,** 223-229*
40. *Zhu C., Gore M., Buckler ES and Yu J 2008 Status and prospects of association mapping in plants. Plant Genome.**1,** 5-20*
41. *Zhu CS andYu JM 2009 Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics.**182,** 875-888*